

GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 00:49:37 ; Search time 2942.31 Seconds
(without alignments)
13738.812 Million cell updates/sec

Title: US-09-768-781-2
Perfect score: 1389
Sequence: 1 atgaacacagaccacaaca.....caaggcaagtggtgtctga 1389

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba.*
- 2: gb_btg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	748.8	53.9	37650	9	HSU131B10	Z73417 Human DNA s
2	622.4	44.8	168918	2	AC116256	AC116256 Rattus no
3	617	44.4	171014	10	AL671915	AL671915 Mouse DNA
4	452.4	32.6	1609	9	AK098608	AK098608 Homo sapi
5	303.4	21.8	2911	10	AF155511	AF155511 Mus muscu
6	291.2	21.0	5096	6	AX332514	AX332514 Sequence
7	291.2	21.0	5096	9	HSXKWP	Z33684 Homo sapien
8	283.6	20.4	5215	6	AX405691	AX405691 Sequence
9	278	20.0	113688	9	AC005301	AC005301 Homo sapi
10	278	20.0	124823	9	AC007064	AC007064 Homo sapi
11	249.8	18.0	168918	2	AC116256	AC116256 Rattus no
12	218	15.7	160200	10	AC091605	AC091605 Mus muscu
13	218	15.7	200698	2	AL672060	AL672060 Mus muscu
14	217.2	15.6	166436	9	AC016752	AC016752 Homo sapi
15	215.6	15.5	161879	9	AC017005	AC017005 Homo sapi
16	215.6	15.5	182083	9	AC007965	AC007965 Homo sapi
17	210.8	15.2	37566	9	AC068541	AC068541 Homo sapi
18	210.8	15.2	174082	9	AC007379	AC007379 Homo sapi
19	207.6	14.9	101912	9	AC000100	AC000100 Homo sapi
20	207.6	14.9	120951	9	AC073649	AC073649 Homo sapi
21	207.6	14.9	145383	9	AC010682	AC010682 Homo sapi
22	207.6	14.9	149794	2	AC007273	AC007273 Homo sapi
23	194.4	14.0	175531	9	HSXKSRPXM	AL121577 Homo sapi
24	179.8	12.9	163495	9	AC024183	AC024183 Homo sapi
25	179.8	12.9	164891	2	AC022848	AC022848 Homo sapi
26	179.8	12.9	195174	9	AC007742	AC007742 Homo sapi
27	173	12.5	1588	9	HSFA000997	AF000997 Homo sapi
28	164	11.8	91640	2	AC123295	AC123295 Rattus no
29	161.8	11.6	626	6	AX079777	AX079777 Sequence
30	104.8	7.5	321	11	G03706	G03706 human STS W
31	91.6	6.5	164891	2	AC022848	AC022848 Homo sapi
32	62	4.5	64450	9	AC073893	AC073893 Homo sapi
33	57.8	4.2	138674	2	AC112084	AC112084 Rattus no
34	51.6	3.7	125020	9	AF429315	AF429315 Homo sapi
35	49.8	3.6	188048	10	AL669974	AL669974 Mouse DNA
36	47.4	3.4	212374	2	AC022773	AC022773 Mus muscu
37	46.8	3.4	68873	2	AC101531	AC101531 Mus muscu
38	46.8	3.4	234491	2	AL671118	AL671118 Mus muscu
39	45.6	3.3	144794	9	AC107055	AC107055 Homo sapi
40	45.4	3.3	7218	6	I66494	I66494 Sequence 14
41	45	3.2	1694	9	AF226053	AF226053 Homo sapi
42	44.4	3.2	13427	6	AX346828	AX346828 Sequence
43	44	3.2	152434	2	AC016262	AC016262 Homo sapi
44	43.8	3.2	104455	2	AC098390	AC098390 Rattus no
45	43.2	3.1	12229	3	AE001377	AE001377 Plasmodiu

ALIGNMENTS

RESULT 1	HSU131B10	DNA	linear	PRI 23-NOV-1999
LOCUS	HSU131B10	37650 bp		
DEFINITION	Human DNA sequence from cosmid U131B10, between markers DXS366 and DXS87 on chromosome X contains XK membrane transport protein, ESTs and STS.			
ACCESSION	Z73417			
VERSION	Z73417.1	GI:1322397		
KEYWORDS	X; XK membrane transport protein.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 37650) Dodsworth, S.			

TITLE Direct Submission
JOURNAL Submitted (17-MAY-1995) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT IMPORTANT: This sequence is the entire insert of clone U131B10.
The true left end of clone U131B10 is at 1 in this sequence. The
true right end of clone U131B10 is at 37650.
U131B10 is from the human chromosome X-specific cosmid library.
FEATURES
Location/Qualifiers
source 1..37650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="LLOXNC01-131B10"
/clone_lib="LLOXNC01"
364..519
/note="MIR element fragment"
repeat_region
2000..2290
/note="MIR element fragment"
repeat_region
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
3347..3637
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
3699..3800
/note="MSTA element fragment"
repeat_region
3712..3842
/note="THE1B element fragment"
repeat_region
3875..4004
/note="MSTA element fragment"
repeat_region
3909..4009
/note="THE1B element fragment"
repeat_region
4038..4085
/note="24 copies of 2 mer 85 & conserved"
repeat_region
4087..4378
/partial
/note="Alu repeat: matches 308. .1 of consensus"
misc_feature
4957..5561
/note="match: Multiple ESTs"
misc_feature
complement(5312..5561)
/note="match: STS G15323"
repeat_region
5723..5842
/note="L1 element fragment"
repeat_region
5980..6292
/note="MLR2C2 element fragment"
repeat_region
5980..6048
/note="MLR2D element fragment"
repeat_region
6399..6759
/note="L1 element fragment"
repeat_region
7148..7437
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
7460..7754
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
9496..9786
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
10339..10630
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
12989..13148
/partial
/note="Alu repeat: matches 308. .145 of consensus"
repeat_region
13166..13450
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
14160..14208
/note="L1 element fragment"
repeat_region
16486..16792
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
17220..17477
/partial
/note="Alu repeat: matches 28. .308 of consensus"
repeat_region
17909..18098

/note="MER3 element fragment"
18505..18795
/note="Alu repeat: matches 1. .308 of consensus"
19300..19436
/partial
/note="Alu repeat: matches 1. .149 of consensus"
19440..19644
/partial
/note="Alu repeat: matches 1. .224 of consensus"
19645..19710
/partial
/note="Alu repeat: matches 242. .308 of consensus"
19711..19870
/partial
/note="Alu repeat: matches 149. .308 of consensus"
21881..22027
/note="MIR element fragment"
23139..23303
/partial
/note="Alu repeat: matches 178. .1 of consensus"
23343..23508
/partial
/note="Alu repeat: matches 308. .138 of consensus"
23509..23800
/partial
/note="Alu repeat: matches 308. .1 of consensus"
23840..23925
/note="MIR2 element fragment"
24408..24890
/note="L1 element fragment"
24527..24554
/note="14 copies of 2 mer 100 & conserved"
24902..24947
/note="23 copies of 2 mer 89 & conserved"
24951..25431
/note="L1 element fragment"
complement(25545..26171)
/partial
/note="match: 232684 XK membrane transport protein"
26196..26539
/note="match: 5' EST H87640 clone 252548; Paired with EST
H87641 matching this cosmid"
complement(26828..27189)
/note="match: 3' EST H87629 clone 252524"
complement(26833..27186)
/note="match: 3' EST H87641 clone 252528; Paired with EST
H87640 matching this cosmid"
27219..27266
/note="24 copies of 2 mer 96 & conserved"
27430..27721
/partial
/note="Alu repeat: matches 308. .1 of consensus"
28995..30018
/note="match: Multiple ESTs"
30042..30333
/partial
/note="Alu repeat: matches 308. .1 of consensus"
30691..30724
/note="17 copies of 2 mer 82 & conserved"
30725..31015
/partial
/note="Alu repeat: matches 308. .1 of consensus"
32221..32499
/note="Alu repeat: matches 1. .301 of consensus"
32500..32529
/note="15 copies of 2 mer 100 & conserved"
33138..33433
/note="Alu repeat: matches 1. .308 of consensus"
33557..33894
/note="MLTID element fragment"
33914..33989
/note="MLTIE element fragment"
33935..34011

repeat_region /note="MLTID element fragment"
34218..34554
/note="L1 element fragment"
35442..35509
/note="MIR element fragment"
36168..36297
/note="L1 element fragment"
36320..36353
/note="17 copies of 2 mer 82 & conserved"
36354..37278
/note="L1 element fragment"
37352..37435
/note="L1 element fragment"
37512..37559
/note="L1 element fragment"
37589..37650
/partial
/note="Alu repeat: matches 1..62 of consensus"
BASE COUNT 10293 a 7831 c 7983 g 11543 t
ORIGIN

Query Match 53.9%; Score 748.8; DB 9; Length 37650;
Best Local Similarity 97.2%; Pred. No. 5.7e-204;
Matches 762; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 606 TGTGAGCTGATCTCTGCAGAGGTTCCCTGGTAGAGTTGCTGCTAAATGTTTCCCT 665
DB 25516 TGATGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 25575
QY 666 GGTATCTGTACCTATGGGCGACCCCTTGGCAATATGTTGGCTATCCAGATCAAGTAGCA 725
DB 25576 GGTATCTGTACCTATGGGCGACCCCTTGGCAATATGTTGGCTATCCAGATCAAGTAGCA 25635
QY 726 TGACTACAAGATTCCCTTGGGCGACCTAGAGTCTCTGCAATCAACATCTGGCGACATT 785
DB 25636 TGACTACAAGATTCCCTTGGGCGACCTAGAGTCTCTGCAATCAACATCTGGCGACATT 25695
QY 786 GGAGATCACTTCCGCTCTGATTTGGTCTCTCTCAGCCATTTGAAATGAAGGC 845
DB 25696 GGAGATCACTTCCGCTCTGATTTGGTCTCTCTCAGCCATTTGAAATGAAGGC 25755
QY 846 TGTGCCCTTCTAGTGTCTCAATCTCTGATCATCTCTTGGAGCCCTGGATTAAGTTCTG 905
DB 25756 TGTGCCCTTCTAGTGTCTCAATCTCTGATCATCTCTTGGAGCCCTGGATTAAGTTCTG 25815
QY 906 GAGAAGTGTGCGCAGATGCCCAATAACATTGAGAAAACTTCAGCCGGGTGCGCACTCT 965
DB 25816 GAGAAGTGTGCGCAGATGCCCAATAACATTGAGAAAACTTCAGCCGGGTGCGCACTCT 25875
QY 966 GGTGCTCTGATTTTCAGTCACATCTCTATGCTGGCATCACTTCTTCTGCTGTCAGC 1025
DB 25876 GGTGCTCTGATTTTCAGTCACATCTCTATGCTGGCATCACTTCTTCTGCTGTCAGC 25935
QY 1026 TTTGAGTGTAGGTTGGCAGACAGATCTCTGTCACAAAGGCGAGAACTGGGACATAT 1085
DB 25936 TTTGAGTGTAGGTTGGCAGACAGATCTCTGTCACAAAGGCGAGAACTGGGACATAT 25995
QY 1086 GGGCCTGCATATATAGTGTGAGGTTGGTAGAAGATGTATGTTCTTGGTTTAAAGTT 1145
DB 25996 GGGCCTGCATATATAGTGTGAGGTTGGTAGAAGATGTATGTTCTTGGTTTAAAGTT 26055
QY 1146 CTTTGGAGTGAAGTGTACTGAATTAATCTGTCATTCCTTGAATGCTGTCAGCTCATAT 1205
DB 26056 CTTTGGAGTGAAGTGTACTGAATTAATCTGTCATTCCTTGAATGCTGTCAGCTCATAT 26115
QY 1206 TGCTATCTGATTTCCATGCTCTCATCTCTTCTTCTCCAGTACTTGCATCCATTCGG 1265
DB 26116 TGCTATCTGATTTCCATGCTCTCATCTCTTCTTCTCCAGTACTTGCATCCATTCGG 26175
QY 1366 CTCACTCTTACCCCAATATGTAGTAGACTACCTCCATTTGTGTCCTGTGTCACGACCCC 1325
DB 26176 CTCACTCTTACCCCAATATGTAGTAGACTACCTCCATTTGTGTCCTGTGTCACGACCCC 26235

QY 1326 TCGGACGAGGTTGAGAACTCAGAGCCACCTTTGAGACTGAAGCAAGCAAGTGTGT 1385
DB 26236 TCGGACGAGGTTGAGAACTCAGAGCCACCTTTGAGACTGAAGCAAGCAAGTGTGT 26295
QY 1386 CTGA 1389
DB 26296 CTGA 26299
RESULT 2
AC116256/168918 bp DNA linear HTG 18-JUL-2002
LOCUS Rattus norvegicus clone CH230-139K19, *** SEQUENCING IN PROGRESS
DEFINITION *** 74 unordered pieces.
AC116256
AC116256.2 GI:21745900
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 168918)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Buck,J., Bowie,S., Brieva,M., Brown,E., Brown,N.P.,
Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Delaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delany,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Dearthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.B.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Maesey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Ogungbe,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Picketts,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanik,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 168918)
Worley,K.C.
Direct Submission
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168918)
Worley,K.C.

TITLE
JOURNAL

Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 14, 2002 this sequence version replaced gi:19718450.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GTLM

Center clone name: CH230-139K19

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 104166 bases at least Q40

Consensus quality: 113489 bases at least Q30

Consensus quality: 119214 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1568: contig of 1568 bp in length
* 1569 1668: gap of unknown length
* 1669 2687: contig of 1019 bp in length
* 2688 2787: gap of unknown length
* 2788 4137: contig of 1350 bp in length
* 4138 4237: gap of unknown length
* 4238 5268: contig of 1031 bp in length
* 5269 5368: gap of unknown length
* 5369 6665: contig of 1297 bp in length
* 6666 6765: gap of unknown length
* 6766 8226: contig of 1461 bp in length
* 8227 8326: gap of unknown length
* 8327 9449: contig of 1123 bp in length
* 9450 9549: gap of unknown length
* 9550 10623: contig of 1074 bp in length
* 10624 10723: gap of unknown length
* 10724 11766: contig of 1043 bp in length
* 11767 11866: gap of unknown length
* 11867 12876: contig of 1010 bp in length
* 12877 12976: gap of unknown length
* 12978 14300: contig of 1324 bp in length
* 14301 14400: gap of unknown length
* 14401 15614: contig of 1214 bp in length
* 15615 15714: gap of unknown length
* 15715 17224: contig of 2010 bp in length
* 17225 17824: gap of unknown length
* 17826 19055: contig of 1231 bp in length
* 19056 19155: gap of unknown length
* 19156 20675: contig of 1520 bp in length
* 20676 20775: gap of unknown length
* 20776 22570: contig of 1795 bp in length
* 22571 22670: gap of unknown length
* 22671 24166: contig of 1496 bp in length
* 24167 24266: gap of unknown length
* 24267 25499: contig of 1233 bp in length
* 25500 25599: gap of unknown length
* 25600 27074: contig of 1475 bp in length
* 27075 27174: gap of unknown length
* 27175 28475: contig of 1301 bp in length
* 28476 28575: gap of unknown length
* 28576 29878: contig of 1303 bp in length
* 29879 29978: gap of unknown length

29979 31798: contig of 1820 bp in length
* 31799 31988: gap of unknown length
* 31989 33456: contig of 1558 bp in length
* 33457 33556: gap of unknown length
* 33557 34753: contig of 1197 bp in length
* 34754 34853: gap of unknown length
* 34854 37077: contig of 2224 bp in length
* 37078 37177: gap of unknown length
* 37178 38599: contig of 1782 bp in length
* 38960 39059: gap of unknown length
* 39060 40603: contig of 1544 bp in length
* 40604 40703: gap of unknown length
* 40704 41879: contig of 1176 bp in length
* 41880 41979: gap of unknown length
* 41980 44025: contig of 2046 bp in length
* 44026 44125: gap of unknown length
* 44126 45351: contig of 1226 bp in length
* 45352 45451: gap of unknown length
* 45452 47009: contig of 1558 bp in length
* 47010 47109: gap of unknown length
* 47110 49436: contig of 2327 bp in length
* 49437 49536: gap of unknown length
* 49537 51265: contig of 1729 bp in length
* 51266 51365: gap of unknown length
* 51366 52885: contig of 1520 bp in length
* 52886 52985: gap of unknown length
* 52986 55166: contig of 2181 bp in length
* 55167 55266: gap of unknown length
* 55267 57181: contig of 1915 bp in length
* 57182 57281: gap of unknown length
* 57282 59138: contig of 1857 bp in length
* 59139 59238: gap of unknown length
* 59239 61142: contig of 1904 bp in length
* 61143 61242: gap of unknown length
* 61243 63220: contig of 1978 bp in length
* 63221 63320: gap of unknown length
* 63321 64522: contig of 1202 bp in length
* 64523 64622: gap of unknown length
* 64623 66069: contig of 1447 bp in length
* 66070 66169: gap of unknown length
* 66170 67304: contig of 1135 bp in length
* 67305 67404: gap of unknown length
* 67405 69305: contig of 1901 bp in length
* 69306 69405: gap of unknown length
* 69406 71453: contig of 2048 bp in length
* 71454 71553: gap of unknown length
* 71554 74540: contig of 2987 bp in length
* 74541 74640: gap of unknown length
* 74641 76351: contig of 1711 bp in length
* 76352 78451: gap of unknown length
* 78452 78453: contig of 2792 bp in length
* 78454 79343: gap of unknown length
* 79344 81674: contig of 2331 bp in length
* 81675 81676: gap of unknown length
* 81677 83496: contig of 1722 bp in length
* 83497 83596: gap of unknown length
* 83597 85142: contig of 1546 bp in length
* 85143 85242: gap of unknown length
* 85243 87769: contig of 2527 bp in length
* 87770 87869: gap of unknown length
* 87870 89602: contig of 1733 bp in length
* 89603 89702: gap of unknown length
* 89703 91131: contig of 1429 bp in length
* 91132 91231: gap of unknown length
* 91232 93807: contig of 2576 bp in length

Query Match 44.8%; Score 622.4; DB 2; Length 168918;

Best Local Similarity 98.6%; Pred. No. 1.9e-167;

Matches 674; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

OY 629 TTCCTTATTTGTCAGTGTGCTAATGGTATTTTCCTGGTATCTGTACCTATGGGGCCA 688

DB 140555 TTCCTTATTTGTCAGTGTGCTAATGGTATTTTCCTGGTATCTGTACCTATGGGGCCA 140496

Db 114791 TCATGCTTCTTTCTTCAGTACTTGCACCCCTGCGCTCACTTCCACCAATACGCTAG 114732
 Qy 1289 TAGACTACCTCCATTGTCTGCTGTCCACGACACCCCTCGACGAGGTTGAGAACTCAG 1348
 Db 114731 TAGACTACCTCCATTGTATGTCTGCGCGACACCCCGTCCGGAAGGTTGAGAACTCAG 114672
 Qy 1349 AGCCACCTTTGAGACTGACGACGACGAGGCAAGTGTGTCTGA 1389
 Db 114671 AGACATCTGTGAAGCTGACACACACAAAGTATTGTCTGA 114631

RESULT 4
 AK098608 1609 bp mRNA linear PRI 15-JUL-2002
 LOCUS Homo sapiens CDNA FLJ25742 fis, clone TST05965.
 DEFINITION AK098608
 ACCESSION AK098608.1 GI:21758666
 VERSION
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens testis cDNA to mRNA, clone lib:TST clone:TST05965.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Ohnima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
 Arita, M., Mutsaers, K., Yuuki, H., Hara, H., Suzuki, Y., Hata, H.,
 Nakagawa, K., Mizuno, S., Morinaga, M., Kawamura, M., Sugiyama, T.,
 Irie, R., Otsuki, T., Sato, H., Nishikawa, T., Sugiyama, A.,
 Kawakami, B., Nagai, K., Isogai, T. and Sugano, S.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1609)
 AUTHORS Sugano, S. and Suzuki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure, Human Genome
 Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction and 5'-end one pass sequencing: Institute of Medical
 Science, University of Tokyo, Laboratory of Genome Structure, Human
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for
 full insert sequencing: RAB and Helix Research Institute.

FEATURES
 source Location/Qualifiers
 1..1609
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="TST05965"
 /tissue_type="testis"
 /clone_lib="TST"
 /note="Cloning vector: pME18SFL3"
 93..1472
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAC05352.1"
 /db_xref="GI:21758667"
 /translation="METVREMDEESTGVSSKEEVLQRLHLSPFSPISITVDYLF
 CGEYVGLGMPFYRKANDTFWSTFISIVGAILDQIILMFNFKDLRRNKAAALFQW
 HILLGLVIRCLARTIRHWKLNKQKEETQVSTPKRNTMLERIAFISIRDNFQWQ
 KAPKMSVIOAFGLSPQVLLQWYISLTIREWPLNRALLMTFSLSVTYGAIRCNILA
 IQSNDDTTKLPIEPFCVVMRFLEVISRVVTLALFIASLKSLPVLIIIVFVSL
 LAPLEWFTFGSAHLPGNKNNMNVGLMFLITLLYALINFSVCWSAVKLQSDDKI
 IDGWRGHRILHYSFQFLNVMILVRFPGGKTLNCCDSLIAVOLIIYSLLATG
 MLFPYQLYLPWQSGKVLPGRTENQPEAPYYVNIETKKNKQLRNCNCRNVGIF
 SIRKSMTCSS"

BASE COUNT 486 a 290 c 336 g 497 t
 ORIGIN

Query Match 32.6%; Score 452.4; DB 9; Length 1609;
 Best Local Similarity 60.7%; Pred. No. 8.8e-119;

Db	581	GTGGAAGAAGTCTCCTCATGACCATATCCCTGTTGTCATTTGTGTATGAGACCTTGGCG	640
Qy	694	TGCAATATGTTGGCTATCCAGATCAAGTACGACTACAAGATTTCGCTTGGGCCACTA	753
Db	641	TGCAACATCCTAGCCATCAAAATCAAGTACGATGAGTATGAGTCAAAGTGAACCTCTG	700
Qy	754	GAAGTCTCTGCATCAGCATCTGGCGGACATTTGAGATCAGTCTCCCGCTCTGATTCGT	813
Db	701	GCCTATGTCGTATCTTTCTGTGGAGGAGCTTTGAGATTTGCCATTCGAGTGTAGTCTG	760
Qy	814	GTGCTCTTCTCAGCGCACTTTGAAATTGAAGCTGTGCCCTTCTCTAGTGCTCAACTTCCGT	873
Db	761	GTCTCTTTTACTCCGCTCGAAGACCTGGGTGGTGTATATATCTATCAACTTCTTCTC	820
Qy	874	ATCATCTCTTTTGAGCCCTCGGATTAAGTTCTTGGAGAAAGTGTGCCAGATGCCCAATAAC	933
Db	821	AGTTTCTTCTGTACCCCTGGATCCTCTTCTGTGTGAGTGGTTCCTCCCATTTCCCTGAGA	880
Qy	934	ATTGAGAAAAAATTCACGCGGGTGGCACTCTGGTGTCTGTATTTTCAGTCACCATCCTC	993
Db	881	ATAGAGAAAGGCCCTCAGTAGAGTGGGACCAACCATTTGTACTATGCTTTTCTAACTTTA	940
Qy	994	TATCTGSCATCAACTTCTCTGTGTGTGAGCTTTGAGTTTGAGTTGGCGACAGAGAT	1053
Db	941	TATACTGATATCAACATGTTCTGTGTGTCTGCTGATCAGCTGAAATTTGACAGCCCTGAC	1000
Qy	1054	CTCGTCGACAAAGGGCAGAACTGGGGACATATGGGCTGCACATATAGTGTGAGTTGGTA	1113
Db	1001	CTCATCAGCAAGTCCCATATTTGGTACCAGCTACTGGTGTATTACATGATAAGATTTCATC	1060
Qy	1114	GAGAAATGTGATCATGCTCTTGGTTTTTAAAGTTCTTTTGGAGTGAAAGTGTACTGAATTAC	1173
Db	1061	GAGAAATGCCATCT	1120
Qy	1174	TGTCATTTCTTGATGCTTGGCAGCTATTAATGCTTATCTGATTTCCATTTGCGCTTCATG	1233
Db	1121	TGCGACCTCTGTTGGTCTGTCAGCTGCTCATTTGGGTACTGTCACAGCCATTTCTTTCT	1180
Qy	1234	CTCCTTTTCTTCCAGTACTTGGCATCCATTTGGCTCACTCTTCCACCCATAATGTAGTAGAC	1293
Db	1181	CTTGATTTCTATCAGTTCTTCCACCTTGCAAAAGCTCTTTTCTTCCAGTGTTCTTGAA	1240
Qy	1294	TACCTCCA 1301	
Db	1241	GGCTTTCA 1248	
RESULT 7			
HSXKMT		5096 bp	linear
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			

AUTHORS	HO,M.F.
TITLE	Direct Submission
JOURNAL	Submitted (21-OCT-1997) Meng F Ho, Human Genetics, Imperial Cancer Research Fund, Institute of Molecular Medicine John Radcliffe Hospital Headington, Oxford, OX3 9DU, United Kingdom
COMMENT	On May 20, 1999 this sequence version replaced gi:2570027.
FEATURES	Location/Qualifiers
source	1. .5096 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="X" /clone="XK" 83. .5068 /gene="XK" 83. .1417 /gene="XK" /citation=[1] /codon_start=1 /product="membrane transport protein" /protein_id="CAA83632.2" /db_xref="gi:4883433" /translation="MKPPASVLASVFLFVAETTAALSLSYTSRGGDMWQALTLIFS LLPCALVQLTLFLVHRDLSDRPLVLLHLGLPLRCFCEVFCIFYQSGNNEPVYS ITKKQMPKGLSEIEKEVQOAEGLITHRSPASRAVIOAFGLGAPQLTLQLYISV MQDVTGVSILMTTISLSIVGALRCNLIAIKIKYDEYVKVKPLAYVCIFLWRSEF TATRVVVLFTSVLTKTWVVVILINFPFLPWLFCWCGSPFPENIEKALSRVGT TIVLCFTLITVLMFCWASQALDKSDPLISKSHNWQLLVYYMIRFIENAILLLL WYLPKTDIYMYVCAPLVLQLLIGYCTAILFLMVLYQFFHPCKKLFSVSEGFQRL RCFCWACRQOKPCPEPIGKEDLQSSRRDRDETPSSSKTSPPEGQFLNAEDLCSA" 83. .327 /gene="XK" /number=1 328. .590 /gene="XK" /number=2 591. .1417 /gene="XK" /number=3 5036. .5041 /gene="XK" 5063. .5068 /gene="XK" 5092 polyA_site 1392 a 1064 c 1022 g 1618 t BASE COUNT ORIGIN

Query Match	21.0%;	Score	291.2;	DB	9;	Length	5096;
Best Local Similarity	54.7%;	Pred.	No. 3e-72;				
Matches	628;	Conservative	0;	Mismatches	508;	Indels	12; Gaps
							2;
Qy	157	TCCACCTTTTGTACTGTGGGAGGCTGCATCTCGCTTTTGTCATGTTTAGAATCTATCGA	216				
Db	110	TCCGTGTTCTCGTTCGCCGAGACAACGGCGCGCTCAGCCTGAGCAGCACCTACCGC	169				
Qy	217	AAGAATAGTGAACCTTACTCGANGACATACACCTTTTCTTCTTTATGTTTTCATTCGATT	276				
Db	170	TCGGCGGGGACC CGCATGTGGCAGCGCTGACGTGCTTTTCTCGCTACTGCCTTGGCG	229				
Qy	277	ATGTCACAGTTGACCCCTCAATTTTTGTCACAGAGATCTAGCCAAAGATAAACCCGCTATCA	336				
Db	230	CTCGTGCAGCTCAGCCTTCTCTCTGTCACCGCACCTCAGCGCGACCGCCCGCTCGTA	289				
Qy	337	TTATTTATGCATCTAACTCTTGGGACCTGTTATCAGATGTTTGGAGGCCATGATTAAAG	396				
Db	290	CTGCTGCTGCACCTGCTGCACTTTGGGCCCCCTTTTCAGGTCTTTTGAAGTCTTCTGCATC	349				
Qy	397	TACCTCACACTGTGGAGAAAAGAGGAGCAGGAGGAGCCCTATGTACGCCCTCACCCGAAG	456				
Db	350	TACTTTC-----AGTCAGGCCAACAAATGAAGAGCCTTATGTCA GTATCACCAAGAAG	400				
Qy	457	AAG----ATGCTAANTAGATGGCGAGGTGCTGATAGAA TGGGAGGTGGGCCACTCCATC	513				
Db	401	AGGCNAATGCCAAAAAATTTGGGCTCTTCAGAGGAGATTGAGAAGGAGGTGGGCCAGGCAGAA	460				

Qy	514	CGGACCTCTGGGTATGCAACCGCAATGCTCAAAACGTATGTCACAGATCAAGCCTTCCTG	573
Db	461	GGCAAACTAATATCAACCAACCATCAGAGGTTTCAGCGGGCTCGGTGATTCATCAGGCTTTCTTG	520
Qy	574	GGCTCAGTGGCCAGCTGACCTATCAGCTCTATGTGAGCGTGTATCTCTGCAGAGTTCC	633
Db	521	GGCTCAGCCCCAGCTGACCCCTACAGCTGTACATTAAGTGTATGCAGCAGACGTCAC	580
Qy	634	CTGGGTAGAGTTGTCTTAATGTAATTTTCCCTGGTATCTGTCAACCTATGGGCGCCACCTT	693
Db	581	GTTGGAAAGAGTCTCTCATGACCATATCCCTGTTGTCATTGTATGAGGCTTGGCG	640
Qy	694	TGCATATGTTGGCTATCCAGATCAAGTACGATGACTACAAGATTCGCTTCGGGCCACTA	753
Db	641	TGCNAACATCTTAGCCATCAAAATCAAGTACGATGAGTATGAAGTCAAAAGTGAAGGCTCTG	700
Qy	754	GAAGTCTCTGCACTCAACATCTGGCGGACATTTGGAGATCACATTCGCGCTCCTGATCTTG	813
Db	701	GCCTATGTCGTATCTCTCTGTGGAGGACTTGGAGTTGCCACTCGAGTTGTAGTCTTG	760
Qy	814	GTGCTCTTCTCAGCGCACTTTGAAATTTGAAGGTGTGCCCTTCTAGTGTCAACTTCCTG	873
Db	761	GTCTCTTTTAACTCCGCTCTGAAGACCTGGGTGGTGGTTATAATACTCATCAACTTCTTC	820
Qy	874	ATCATCTCTTTGAGCCCTGGATTAAGTTCTGGAGAGTGGTCCCGAGATGCCCAATAC	933
Db	821	AGTTTCTTCTTGTAACCCCTGGATCTCTTCTGTGTGAGTGGTTCGCCCATTCCTCTGAGAAC	880
Qy	934	ATTGAGAAAAACTTCAGCGGGTCGGCACTCTGGTGGTCCGATTTTCAGTCACCACTCTTC	993
Db	881	ATAGAGAGGCCCTCAGTAGAGTGGSCACCACTGTACTATGCTTTCTTACTTTACTTC	940
Qy	994	TATGCTGGCATCAACTCTCTTGCTGGTCAGCTTTGCAGTTGAGGTTGGCAGACAGAGAT	1053
Db	941	TATACTGGTATCAACATGTTCTGCTGGTCTGCTGTACAGCTGAAAAATTGACAGCCCTGCAC	1000
Qy	1054	CTCGTCGACAAAGGCGAGAACTGGGGACATATGGGCTCGCACTATAGTGTGAGGTTGGTA	1113
Db	1001	CTCATCATGCAAGTCCCATTAATTGGTACAGCTACTTGGTGTATTTACATGATAGATTCTATC	1060
Qy	1114	GAGATGTGATCATGCTCTGGTTTTTAAGTTCTTTGGAGTGAAGTGTACTGTAATAC	1173
Db	1061	GAGAAATGCCATCTCTCTCTCTGTGGTATCTTTTCAAGACTGCATCTATATGTATGTG	1120
Qy	1174	TGTCATTTCTTGATTGCCCTGGAGCTCATTTATGCTTATCTGATTTCCATTGGCTTCATG	1233
Db	1121	TGCGACCTCTGTGTGTCTCTGAGCTGCTCATTTGGGTACTGACAGCCATTCTCTTCATG	1180
Qy	1234	CTCTTTTCTTCCAGTACTTGGATCCATTGCGCTCACTCTTCACCCCATTAATGTAGTAGAC	1293
Db	1181	CTTGATTTCTATCAGTTCTTCCACCCTTGCAAAAGCTCTTTTCTTCCAGTGTTTCTGAA	1240
Qy	1294	TACCTCA 1301	
Db	1241	GGCTTTCA 1248	

RESULT 8					
AX405691					
LOCUS		5215 bp	DNA	linear	PAT 14-JUN-2002
DEFINITION	AX405691				
Sequence 106 from Patent WO0222660.					
ACCESSION	AX405691				
VERSION	AX405691.1	GI:21438821			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.					
REFERENCE	1				
AUTHORS	Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F., Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.				
TITLE	Novel nucleic acids and polypeptides				

Qy	235	TGATGACATACACCTTTTCTTCTTATATGTTTTCATTCATTATGGTCAGTTGACCCCTC	294
Db	307	TGGCAGCGCTGACGTTGCTTTTCTCGCTACTGCTTTCGCGCTGTCAGCTCAGCTT	366
Qy	295	ATTTTGTCCACAGAGATCTAGCCAAAGATAAACCGCTATCATTTATGATCATTAATC	354
Db	367	CTCTTCGTACACGCGACCTCAGCCGCGACCGCCGCTCGTACTGCTGTCACCTGCTG	426
Qy	355	CTCTTGGGACCTGTATCAGATGTTTGGAGGCCATGATTAAGTACTCTACACTGTGGAG	414
Db	427	CAACTTGGGCCCCCTTTTCAGGTGTTTGAAGTCTTCTGCACTACTTTC-----AG	477
Qy	415	AAAGAGGAGGAGGAGGCGCTATGTGAGCCTCACCCGAAAGAAG---ATGCTAATAGAT	471
Db	478	TCAGGCAACAAATGAAGAGCCTTATGTCAGTATCACCAAGAAGAGGCAAAATGCCAAAAT	537
Qy	472	GGCGAGGAGGTGCTGATAGAAATGGAGAGGTGGGCCATCTCCATTCGGACCCCTGGCTATGCAC	531
Db	538	GGCCTCTCAGAGGAGATGAGAGAGAGGTGGGCGAGGAGGCAAGAGGCAAACTAATCACCAC	597
Qy	532	CGCAATGCCCTACAAACGTAATGTCACAGATCCAAAGCCTCTCGGGCTCAGTGCCCGAGTG	591
Db	598	CGATCAGGTTTACGCGGGGCTCGGTGATCCAGGCTTTCTTGGGCTCAGCCCCCAGCTG	657
Qy	592	ACCTATCAGCTCTATGTGAGCCTGATCTCTGACAGAGTTTCCCTGGGTAGAGTTGTGCTA	651
Db	658	ACCTACAGCTGTACATAGTGTCTATGACAGCAGACGTCACTGTTTGGAAAGAGTCTCCTC	717
Qy	652	ATGGTATTTTCCCTGGTATCTGTCACTATAGGGGCCACCTTTTGCAAATATGTTGCTATC	711
Db	718	ATGACCATATCCCTGTTGTCCATTGTGTATGAGAGCCTTGGCGCTGCAACATCTAGCCATC	777
Qy	712	CAGATCAAGTACGATCACTACAAAGATTCCGCTTGGGCCACTAGAAGTCTCTGCAATCAC	771
Db	778	AAATCAAGTACGATGAGTATGAAGTCAAAGTGAAGCCCTCTGGCCATGTCTGTGATCTTC	837
Qy	772	ATCTGGCGGACATTGAGAGATCACTTCCCGCCTCCTGATTTCTGGTGTCTTCTCAGCCACT	831
Db	838	CTGTGGAGAGCTTTGAGATTTGCCATCTCAGATTGTAGTCTCTGGTCTCTTTTACCTCCGTC	897
Qy	832	TTGAAATTTGAAGGCTGTGCCCTTCTTAGTGTCTCAACTTCTCGATCATCTCTTTTGGAGCC	891
Db	898	CTGAAGACCTGGTGGTGTATTAATACTCATCAACTTCTTCAGTTTCTCTTGTACCCC	957
Qy	892	TGATTTAAGTCTCGGAAGAGTGGTGGCCAGATGCCCAATAACATTTGAGAAAAAATTCAGC	951
Db	958	TGATCCTCTTCTGGTGCAGTGGTTCCTCATTTCCCTGAGAACATAGAAGGCCCTCAGT	1017

there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below.
AC005301(p15j16) 77414 113688 (0) overlaps AC007064(p8708) 1 36275 (88548).

```

FEATURES             Location/Qualifiers
     source            1..113688
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="22q11"
                        /clone="p15j16"
BASE COUNT           37158 a 22385 c 21889 g 32256 t
ORIGIN
Query Match          20.0%; Score 278; DB 9; Length 113688;
Best Local Similarity 65.1%; Pred. No. 3.1e-68;
Matches 410; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 640 AGAGTTGCTGAATGTTTCCCTGGTATCTGTACCACTATGGGGCCACCTTTGGCAAT 699
Db 101529 ATAGCAATTCGTGATGACATATTTCCCTGTTATCAGTTACTTATGGGGCCATTCGCTGCAAT 101470

QY 700 ATGTGGCTATCCAGATCAAGTACGATGACTACAAAGATTTCGCTTTGGGGCCACTAGAGTC 759
Db 101469 ATACTGGCCATCCAGATCAGCAATGATGATATCTACATTAAGCTACCGCCGATAGAATTC 101410

QY 760 CTCTGSCATCACCATCTGGCGGACATTTGGAGATCACTTCCCGCTCTCTGATTCCTGGTGCTC 819
Db 101409 TCTGTGTCGTGATGCGGCTTTTGGAGGTATCTACGTGTAGTGACTTCGCAATTT 101350

QY 820 TTCTCAGCCACTTTGAATGAAGCTGTGCCCTTCTTAGTGCTCAACTTCTCGATCATC 879
Db 101349 TTCATTGCATCTCTGAACCTGAAGAGCTACCCGTTTGTGTTAATCATATATTTCGTATCA 101290

QY 880 CTCCTTGAGCCCTGATTAAGTTCCTGGAGAAGTGTGGCCAGATGCCCAATCAATTTGAG 939
Db 101289 TTGTTGGCACCCTGGCTGGAGTTTGGAAAAGTGGAGCTCATCTTCTCGGCAACAAGAA 101230

QY 940 AAAAACTTCAGCCGGTTCGGCACTCTGTGGTCTGATTTCAGTCACCATCTCTATGCT 999
Db 101229 AATAATTCATATATGGTGGTACAGTACTGATCTTTCTTGATCACACTGCTATATGCT 101170

QY 1000 GGCATCAACTTCTCTGTCTGTGCTAGCTTTGCAGTTGAGGTTGGCAGACAGAGATCTCGTC 1059
Db 101169 GCCATCAACTTCTCTGTCTGTGCTAGCACTGAACCTGCGATTGTCAGATGACAAATAAT 101110

QY 1060 GACAAAGGGCAGAACTGGGGCAGATATGGCCCTGCACTATAGTGTGAGGTTGTAGAGAA 1119
Db 101109 GACGGGACAGAGGTGGGGCCATAGAATCCTACATCAGCTTTCAGTTTTTAGAAAAT 101050

QY 1120 GTGATCATGGTCTTGGTTTTTAAGTTCTTTGGAGTGAAGTTTACTGAATTAAGTCTCAT 1179
Db 101049 GTGATTAATGATATTTGTTATTTAGGTTCTTTGGAGGAAACCTTCTGCTGAATTTGTGTAC 100990

QY 1180 TCCTTGATGCTTGCAGCTCATTTATGCTTATCTGATTTTCATTTGCTTTCATGCTCTCTT 1239
Db 100989 TCATTAATGCGGTGCAGCTCATCATAGCTACCTATTGGCCACTGGCTTTATGCTCTCTC 100930

QY 1240 TTCTTCAGTACTTGCATCCATTTGGGCTCA 1269
Db 100929 TTCTATCAGTATTTGTACCCATGGCAGTCA 100900

RESULT 10
AC007064/c          124823 bp  DNA  linear  PRI 27-MAY-2000
LOCUS              Homo sapiens Chromosome 22q11 PAC Clone p8708 in CES region,
DEFINITION          complete sequence.
ACCESSION          AC007064
VERSION            AC007064..27 GI:7109507
KEYWORDS           HTG.
SOURCE             Homo sapiens.
ORGANISM           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE

1 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Homo sapiens Chromosome 22q11 PAC Clone p8708 in CES Region
unpublished

REFERENCE

2 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (11-MAR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

3 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (26-OCT-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

4 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (27-OCT-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

5 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (05-NOV-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

6 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (08-NOV-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

7 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (29-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

8 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (25-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE

9 (bases 1 to 124823)
Ray, L.A., Loh, P., Qi, S., Sloan, D., McDermid, H. and Roe, B.A.
Direct Submission
Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT

On Feb 29, 2000 this sequence version replaced gi:6249691.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC005301(p15j16) 77414 113688 (0) overlaps AC007064(p8708) 1 36275
(88548) AC007064(p8708) 62616 124823 (0) overlaps AC006548(p20k14)
1 62199 (137950).

FEATURES

Location/Qualifiers
1..124823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q11"
/clone="p8708"

BASE COUNT 41074 a 23932 c 23324 g 36493 t

ORIGIN

Query Match 20.0%; Score 278; DB 9; Length 124823;
Best Local Similarity 65.1%; Pred. No. 3.2e-68;
Matches 410; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 640 AGAGTTGGCTAATGGTATTTTCCTGGTATCTGTCACTATGGGGCCACCCTTTGCAAT 699
 DB 24116 ATAGCATTTGCTGATGACATTTTTTCCCTGTTATCATGTTACTTATGGGGCCATTCGCTGCAAT 24057
 QY 700 ATGTTGGCTATCCAGATCAAGTACATCAAGATTCGGCTTGGGGCCACTAGAAGTC 759
 DB 24056 ATACTGGCCATCCAGATCAGCAATGATGATACATTAAGCTACCCCGGATAGAAATTC 23997
 QY 760 CTCTGTCATCACCATCTGGCGGACATTTGAGATCACTTCCCGCTCTCTGATTTCTGGTGCTC 819
 DB 23996 TTCTGTGCTGATGCTGGCGTTTTTGGAGGTTATCTCAGCTGTAGTACCTGCGCATTT 23997
 QY 820 TTCTCAGCCACTTTGAAATGAAGCTGTGCCCTTCTTAGTGTCTCAACTTCTCTGATCATC 879
 DB 23936 TTCAATGTCATCTCTGAACTGAAGAGCTACCCGTTTTGTTAATCATATATTTTGTATCA 23877
 QY 880 CTCTTTGAGCCCTGGATTAAGTTCTGGAGAAGTGTGGCCAGATGCCCAATCAACTTGAAG 939
 DB 23876 TTGTGTGACCCGCTGGCTGGAGTTTTTGGAAAAAGTGGAGCTCATCTTCTGGCAACAAGAA 23817
 QY 940 AAAAATTCAGCCGGTGGCACTCTGGTGGTCTGATTTTCAGTCACCATCTCTATGCT 999
 DB 23816 AATAATTTCCATATGTTGGGTACAGTACTGATCTTTCTTGTATCAGTCTGCTATATGCT 23757
 QY 1000 GGCATCAACTTCTCTTGTGCTGCTGAGCTTTTGCAGTTGAGGTTGGCAGACAGATCTCGTC 1059
 DB 23756 GCCATCAACTTCTCTGCTGCTGAGCTGAACTGCAGTTGTGATGACAAATAATTT 23697
 QY 1060 GACAAAGGCGAGAACTGGGACATATGGGCTGCACTATAGTGTGAGGTTGGTAGAAT 1119
 DB 23696 GACGGGACAGAGTGGGGCCATAGAACTCTACACTACAGCTTTCAGTTTTTAGAAAAT 23637
 QY 1120 GTGATCATGTTGTTGTTTAACTTCTTGGAGTGAAGTGTACTGATTTACTGATCTCAT 1179
 DB 23636 GTGATAATGATATGTTATTTAGTTTCTTGGAGGAAACTTCTGCTGAATGTTGTGAC 23577
 QY 1180 TCCTTTGCTTGGCTTGCAGTCAATATGCTATCTGATTTCCATTTGGCTTCACTGCTCTT 1239
 DB 23576 TCATTAATGGCGTGCAGTCACTATAGTACCTATTGGCCACTGGCTTTATGCTCTC 23517
 QY 1240 TTCTTCCAGTATGTCATPCCATTTGGCTCA 1269
 DB 23516 TTCTATCAGTATTGTATCCCATGGCAGTCA 23487
 RESULT 11
 AC116256
 LOCUS Rattus norvegicus clone CH230-139K19, *** SEQUENCING IN PROGRESS
 DEFINITION *** 74 unordered pieces.
 AC116256
 VERSION AC116256.2 GI:21745900
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 168918)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alebrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Deigado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
 Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,N., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
 Ogrunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savary,G.,
 Scher,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
 Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinsan,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 168918)
 Worley,K.C.
 Direct Submission
 Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 168918)
 Worley,K.C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:19718450.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GTLM
 Center clone name: CH230-139K19
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 104166 bases at least Q40
 Consensus quality: 113489 bases at least Q30
 Consensus quality: 119214 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1568: contig of 1568 bp in length
 * 1569 1668: gap of unknown length
 * 1669 2687: contig of 1019 bp in length
 * 2688 2787: gap of unknown length
 * 2788 4137: contig of 1350 bp in length
 * 4138 4237: gap of unknown length

* 4238 5268: contig of 1031 bp in length
* 5269 gap of unknown length
* 5369 contig of 1297 bp in length
* 6665 gap of unknown length
* 6666 gap of unknown length
* 6766 contig of 1461 bp in length
* 8227 gap of unknown length
* 8327 contig of 1123 bp in length
* 9449 gap of unknown length
* 9450 gap of unknown length
* 9550 contig of 1074 bp in length
* 10624 gap of unknown length
* 10724 contig of 1043 bp in length
* 11767 gap of unknown length
* 11867 contig of 1010 bp in length
* 12877 gap of unknown length
* 12977 contig of 1324 bp in length
* 14301 gap of unknown length
* 14401 contig of 1214 bp in length
* 15614 gap of unknown length
* 15615 contig of 2010 bp in length
* 15715 gap of unknown length
* 17725 gap of unknown length
* 17824 contig of 1231 bp in length
* 19085 gap of unknown length
* 19056 gap of unknown length
* 19156 contig of 1520 bp in length
* 20676 gap of unknown length
* 20776 contig of 1795 bp in length
* 22571 gap of unknown length
* 22571 contig of 1496 bp in length
* 24166 gap of unknown length
* 24167 contig of 1233 bp in length
* 24267 gap of unknown length
* 25499 gap of unknown length
* 25500 contig of 1475 bp in length
* 25600 gap of unknown length
* 27074 contig of 1301 bp in length
* 27175 gap of unknown length
* 28476 gap of unknown length
* 28575 contig of 1303 bp in length
* 29878 gap of unknown length
* 29879 contig of 1820 bp in length
* 29979 gap of unknown length
* 31799 gap of unknown length
* 31899 contig of 1558 bp in length
* 33457 gap of unknown length
* 33557 contig of 1197 bp in length
* 34754 gap of unknown length
* 34853 contig of 2224 bp in length
* 34854 gap of unknown length
* 37078 contig of 1782 bp in length
* 37177 gap of unknown length
* 38959 contig of 1782 bp in length
* 39059 gap of unknown length
* 39060 contig of 1544 bp in length
* 40603 gap of unknown length
* 40604 contig of 1176 bp in length
* 40704 gap of unknown length
* 41880 contig of 2046 bp in length
* 41980 gap of unknown length
* 44026 gap of unknown length
* 44125 contig of 1226 bp in length
* 45351 gap of unknown length
* 45352 contig of 1558 bp in length
* 45452 gap of unknown length
* 47010 contig of 2327 bp in length
* 47110 gap of unknown length
* 49437 contig of 1729 bp in length
* 49537 gap of unknown length
* 51265 gap of unknown length
* 51366 contig of 1520 bp in length
* 52886 gap of unknown length
* 52986 contig of 2181 bp in length
* 55166 gap of unknown length
* 55266 contig of 1915 bp in length
* 55267 gap of unknown length
* 57181 contig of 1857 bp in length
* 57282 gap of unknown length
* 59138 contig of 1904 bp in length
* 59239 gap of unknown length
* 61142 gap of unknown length
* 61243 contig of 1978 bp in length
* 63220 gap of unknown length
* 63221 contig of 1202 bp in length
* 64522 contig of 1202 bp in length

* 64523 64622: gap of unknown length
* 64623 contig of 1447 bp in length
* 66070 gap of unknown length
* 66170 contig of 1135 bp in length
* 67305 gap of unknown length
* 67405 contig of 1901 bp in length
* 69306 gap of unknown length
* 69406 contig of 2048 bp in length
* 71453 gap of unknown length
* 71553 contig of 2987 bp in length
* 71554 gap of unknown length
* 74541 gap of unknown length
* 74641 contig of 1711 bp in length
* 76352 gap of unknown length
* 76452 contig of 2792 bp in length
* 79244 gap of unknown length
* 79344 contig of 2331 bp in length
* 81675 gap of unknown length
* 81775 contig of 1722 bp in length
* 83497 gap of unknown length
* 83597 contig of 1546 bp in length
* 85143 gap of unknown length
* 85243 contig of 2527 bp in length
* 87769 gap of unknown length
* 87770 contig of 1733 bp in length
* 88702 gap of unknown length
* 89603 contig of 1429 bp in length
* 89703 gap of unknown length
* 91132 contig of 2576 bp in length
* 91232 gap of unknown length
* 91332 contig of 2576 bp in length
* 91332 contig of 2576 bp in length

Query Match 18.0%; Score 249.8; DB 2; Length 168918;
Best Local Similarity 93.9%; Pred. No. 4.5e-60;
Matches 260; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 372 CAGATGTTTGGAGGCCATGATTAACTACCTCAGCTGTGGAGAAAGAGGAGGAGGAGA 431
Db 122642 CAGATGTTTGGAGGCCATGATTAACTACCTCAGCTGTGGAGAAAGAGGAGGAGGAGA 122701
Qy 432 GCCCTATGTCAGCCTCACCCGAAAGAGATGCTAATAGATGCGAGGAGGCTCTGATAGA 491
Db 122702 GCCATATGTCAGCCTCACCCGAAAGAGATGCTAATAGATGCGAGGAGGCTCTGATAGA 122761
Qy 492 ATGGGAGGTGGGCCATCCATCCGACCCCTGGCTATGACCCGCAATGCTACAAACGTAT 551
Db 122762 ATGGGAGGTGGGCCATCCATCCGACCCCTGGCTATGACCCGCAATGCTACAAACGTAT 122821
Qy 552 GTCACAGATCCAGCCTTCTGGGCTAGTCCGCCAGCTGACCTATCAGCTTATGTGTAG 611
Db 122822 GTCACAGATCCAGCCTTCTGGGCTAGTCCGCCAGCTGACCTATCAGCTTATGTGTAG 122881
Qy 612 CCTGATCTCTGAGAGGTCCCTGGGTAGAGTTGTG 648
Db 122882 CCTGATCTCTGAGAGGTCCCTGGGTAGAGTTGTG 122918

RESULT 12
AC091605/c
LOCUS AC091605 160200 bp DNA linear ROD 02-AUG-2002
DEFINITION Mus musculus Strain C57BL6/J Chromosome X BAC, RP23-64A09, Complete
Sequence, complete sequence.
AC091605
ACCESSION AC091605.6 GI:22091329
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 160200)
AUTHORS Montgomery, K.T., Grills, G., Li, L., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
TITLE High Throughput Mouse Sequencing
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160200)

ORIGIN

Query Match	15.7%	Score 218;	DB 2;	Length 200698;
Best Local Similarity	57.5%;	Pred. No. 6.7e-51;		
Matches 392;	Conservative 0;	Mismatches 290;	Indels 0;	Gaps 0;
QY 605	ATGTGAGCGCTCATCTCTCGACAGAGGTTC	CCCTGGGTTAGAGTTGTGCTAATGGTATTTTCCC	664	
DB 119410	AGGTGAGATTGACAGTGTCTTTCTCTCTCTT	CAGGCTTCATCATGACCTTGTCTCC	119469	
QY 665	TGGTATCTGTCACTATGGGCCACCCCTTTG	CAATATATGTGGCTATCCAGATCAAGTACG	724	
DB 119470	TGCTGTCTATTGTCTATGCGCCCTTACGTT	TGCAACATCTTAGCCATCAAAATCAAGTATG	119529	
QY 725	ATGACTACAAATTCGCCCTTGGGCACCTAG	AAGTCTCTGCATCACCATCTGGGGACAT	784	
DB 119530	ATGAGTATGAGGTCAAAGTCAAACCCCTGG	GTATGTCTGATCTCTCTCTGGAGAACT	119589	
QY 785	TGGAGATCACATTCGCCGCTCTCGATTCT	TGGTGTCTCTTCAGCCACTTTGAAATTGAAGG	844	
DB 119590	TTGAGATTGCCACCGAGTCATCGTCTGGT	CCCTCTTACCTCTGCTCGAAGATCTGGG	119649	
QY 845	CTGTGCCCTTCTTAGTGTCTCAACTTCTCT	GTGATCATCTCTTTGAGCCCTTGGATTAA	904	
DB 119650	TGGTGGCAGTCATACTCGTCAACTTCTT	CAGCTTCTTATATCCCTGGATCGTTTCT	119709	
QY 905	GGAGAAGTGTGGCCAGATGCCCAATAACAT	TGAGAAAAAATTCAGCCGGTGGGCATC	964	
DB 119710	GGTCAGTGGCTCCCCCATTCCTCGAGAA	CATAGAGAAGCCCTAAGTAGGGTGGGTACCA	119769	
QY 965	TGSGTGCTCTGATTCTTCAGTCAACATCCT	CTATGCTGGCATCAACTTCTTGTGCTGTG	1024	
DB 119770	CCATCGTCTCTGCTTCTCTCACTTTACT	TCTATGCGCGCATCAACATGTTCTGCTGT	119829	
QY 1025	CTTTGCAGTTGAGGTTGGCAGACAGATCT	CTGTCGACAAAGGGCAGAACTGGGGACATA	1084	
DB 119830	CTGTTCACTGCTGAAATCGACAAATCGG	AGCTCATCAGCAAGTCCCGAAGCTGATACG	119889	
QY 1085	TGGCCCTGCACCTATAGTGTGAGGTTGGT	TAGAGAAATGTGATCATGGCTCTGGTTTTA	1144	
DB 119890	TGCTCATTTATACATGACGAGATTCAT	TCGAGAACTCCGTCCTCTGCTCTGTGTTAT	119949	
QY 1145	TCTTTGGAGTGAAGTGTTACTGGAATTACT	GTCAATTCCTTCGATTCGCTTCGAGCTCAT	1204	
DB 119950	TTTTTAAAACTGACATATACATGATGTGT	GTGGCCCTCTGTGATCTTCGAGCTGCTT	120009	
QY 1205	TTGCTTATCTGATTTCCATTGGCTTCAT	GTCTCCCTTTCTCCAGTACTTGCATCCATT	1264	
DB 120010	TTGGGTACTGCACGTGGCATCTCTTCAT	GTCTGTGTTCTATCAGTGTTCACCCCTT	120069	
QY 1265	GCTCACTCTTCCACCCATAATGT	1286		
DB 120070	AAAAGCTCTTCTCTCCAGTGT	120091		

[illegible]

AUTHORS TITLE JOURNAL REFERENCE	AUTHORS TITLE JOURNAL REFERENCE	AUTHORS TITLE JOURNAL REFERENCE	AUTHORS TITLE JOURNAL REFERENCE	AUTHORS TITLE JOURNAL REFERENCE
--	--	--	--	--

Nguyen,C., Maupin,R., Hawkins,M. and Smith,R.
The sequence of Homo sapiens BAC clone RP11-506M9
Unpublished
3 (bases 1 to 166436)
Waterston,R.H.
Direct Submission
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 166436)
Waterston,R.H.
Direct Submission
Submitted (23-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 166436)
Waterston,R.H.
Direct Submission
Submitted (24-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 166436)
Waterston,R.H.
Direct Submission
Submitted (25-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
7 (bases 1 to 166436)
Waterston,R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 23, 2000 this sequence version replaced gi:6524293.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H NH0506M09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

MAPPING INFORMATION:
The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

Source: The RFL-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P. Y., Zhao, S., Fritngen, E., Taten, M., Caninese, J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) (VECTOR: pBAC3.6)

NEIGHBORING SEQUENCE INFORMATION:

This clone sequenced to the left is RP11-221K4. Actual start of this clone is at base position 1 of RP11-506M9; actual end is at base position 166436 of RP11-506M9.

FEATURES

source
Location/Qualifiers
1. 166436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-506M9"
/clone_lib="RPCI-11"
195. 923
/rpt_family="L1"
924. 1077
/rpt_family="L1"
1080. 1830
/rpt_family="L1"
1879. 1918
/rpt_family="L1"
1905. 3257
/rpt_family="L1"
3280. 3393
/rpt_family="Alu"
3618. 3694
/rpt_family="MIR"
4250. 4511
/rpt_family="Alu"
4512. 5804
/rpt_family="L1"
5804. 7682
/rpt_family="L1"
8188. 8294
/rpt_family="L2"
9429. 9499
/rpt_family="L2"
9826. 9872
/rpt_family="(A)n"
9962. 10035
/rpt_family="L2"
10596. 10684
/rpt_family="GA-rich"
12836. 13209
/rpt_family="MaLR"
14667. 14689
/rpt_family="AT-rich"
16692. 16712
/rpt_family="AT-rich"
17516. 17584
/rpt_family="MERS3"
17590. 17667
/rpt_family="MERS3"
17814. 18112
/rpt_family="Alu"
19421. 19650
/rpt_family="L1"
19689. 19813
/rpt_family="Retroviral"
19815. 20107
/rpt_family="Retroviral"
20410. 20861
/rpt_family="Retroviral"
21083. 25178
/rpt_family="Retroviral"
25196. 27960
/rpt_family="Retroviral"
28006. 29520
/rpt_family="Retroviral"
28828. 28905
/rpt_family="L2"
29157. 29395
/rpt_family="Alu"
29396. 29438

repeat_region
/rpt_family="(TAA)n"
30965. 30996
/rpt_family="AT-rich"
31776. 32149
/rpt_family="MaLR"
33205. 33489
/rpt_family="Alu"
35802. 35832
/rpt_family="AT-rich"
36311. 36389
/rpt_family="(TTAAA)n"
36757. 36827
/rpt_family="MERS3"
38147. 38551
/rpt_family="Retroviral"
38552. 38576
/rpt_family="(TG)n"
38577. 38633
/rpt_family="Retroviral"
38634. 39399
/rpt_family="Retroviral"
39400. 39459
/rpt_family="MER1_type"
39460. 40129
/rpt_family="Retroviral"
40121. 41902
/rpt_family="Retroviral"
41931. 42280
/rpt_family="Retroviral"
42281. 42759
/rpt_family="Retroviral"
42961. 43305
/rpt_family="L1"
43410. 43703
/rpt_family="Alu"
44044. 44848
/rpt_family="L1"
44853. 45060
repeat_region
Query Match 15.6%; Score 217.2; DB 9; Length 166436;
Best Local Similarity 62.3%; Pred. No. 1.1e-50;
Matches 391; Conservative 0; Mismatches 233; Indels 4; Gaps 3;
Qy 642 AGTTGTCTAATGGTATTTCCTGGTATCTGTACCTATGGGCCACCTTTGGCAATAT 701
Db 114404 AGCATTCGTGATGACATGTTCCCTCTTACCACCTGTGTATGGGCCACTTCACTGCAATAC 114345
Qy 702 GTTGGCTATCCAGATCAAGTACGATGACTACAAGATTCGCTTGGGCCACTAGAAAGTCT 761
Db 114344 ACTGGCCATCCATACAGCAGCAGCTTCTCGATTAAGCTACAGCGATAAAATTCAT 114285
Qy 762 CTGCATCACCATCTGGCGGACATTTGGAGATCATTCCCGCCTCCTGTATTTGGTCTCTT 821
Db 114284 CTG--TCATGATACGGTGTCTCTTGGTGATTATCTCAGCTAGTAGCTCTGGAATTTT 114227
Qy 822 CTGAGCCACTTTGAAATGAAGCTGTGCCCTTCTTAGTGTCTCAACTTCCTGATCATCT 881
Db 114226 TCCTGCATCTCTGAACAGAGGAGCCTACCTTACTATTATATATATTTTGTATTAT 114167
Qy 882 CTTTGAGCCCTGGATTAAAGTTCTGGAGAAGTGGTGCCAGATGCCAATAACATTGAGAA 941
Db 114166 GTTGGCACCATGGCTGGAGTTTGGAAAGTGGAGCTCATCTTATAACAACACAGAAA 114107
Qy 942 AAACCTTCAGCCGGTCCGCACTCTGGTGGTCTGTGATTTCACTACCATCTCTATGCTGG 1001
Db 114106 TAATTCAGCATGGTGGGTAC-ATACGATGCTTACCTTAATCATGTCTACTATATGCTG 114048
Qy 1002 CATCAACTTCTTGTGGTGCAGCTTTGAGGTTGGCAGACAGAGATCTCGTGA 1061
Db 114047 TATCAACTTCTTGTGGTGGTGCAGCACTGAAACTGCAGCTCTCAAATGAGGAATTA 113988
Qy 1062 CAAAGGCGAGAACTGGGGACATATGGGCCCTGCATATAGTGTAGGTTGGTAGAATGT 1121

Db 113987 CAAGAGACTGAGGTGGC-CATAGAATCCTACACTACAGCTTCTAGTTTTTTAGAAAAATGT 113929

Qy 1122 GATCATGGTCTTGGTTTTTAAGTCTCTTTGGAGTGAAGTGTACTGAATTACTGTGTCATTC 1181

Db 113928 GATAATAATATTGATATTATGTTCTGTTGAGAGGAAATTTCACTGAAGCTTTGTGACTC 113869

Qy 1182 CTTGATTGCTTGCAGCTCAATTATTGCTTATCTGATTTCCATTTGGCTTCATGCTCTTTT 1241

Db 113868 ATTAATTGCCATGTTAGTTCATCATAACTCACTAATATTAGCAATTTGGCTTTATGCTCTCTCTT 113809

Qy 1242 CTTCCAGTACTGTCATCCATTCGGCTCA 1269

Db 113808 CTGTCAAGTATTTCACCCCAAGGTGTCA 113781

RESULT 15

LOCUS AC017005 161879 bp DNA linear PRI 09-MAY-2001

DEFINITION Homo sapiens BAC clone RP11-100J21 from Y, complete sequence.

ACCESSION AC017005

VERSION AC017005.6 GI:13435269

KEYWORDS HTG.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 161879)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

PUBMED 9847074

REFERENCE 2 (bases 1 to 161879)

Armstrong, J., Stoneking, T., Hawkins, M. and Bernard, K.

The sequence of Homo sapiens BAC clone RP11-100J21

Unpublished

REFERENCE 3 (bases 1 to 161879)

Waterston, R.H.

Direct Submission

Submitted (09-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 161879)

Waterston, R.H.

Direct Submission

Submitted (23-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 161879)

Waterston, R.

Direct Submission

Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Mar 23, 2001 this sequence version replaced gi:19211335.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

Center project name: H_NH0100J21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-79J10; the clone sequenced to the right is RP11-245K4, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-100J21; actual end is at base position 66964 of RP11-245K4.

FEATURES

source	Location/Qualifiers
1. .161879	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="Y"
	/map="Y"
	/clone_lib="RPCI-11"
1. .1185	/rpt_family="L1"
1183. .1513	/rpt_family="L1"
1679. .1842	/rpt_family="MER1_type"
2125. .3135	/rpt_family="ERV1"
3116. .3412	/rpt_family="ERV1"
3450. .3866	/rpt_family="ERV1"
3867. .6977	/rpt_family="ERV1"
6986. .7295	/rpt_family="ERV1"
7297. .7625	/rpt_family="ERV1"
9507. .9570	/rpt_family="ERV1"
11143. .11817	/rpt_family="MER1_type"
12324. .12843	/rpt_family="L1"
12874. .12922	/rpt_family="L1"
12885. .12929	/rpt_family="ERV1"
13051. .13264	/rpt_family="ERV1"
13381. .13619	/rpt_family="L1"
13628. .14306	/rpt_family="L1"
15532. .15662	/rpt_family="ERV1"
15740. .16044	/rpt_family="ERV1"
16045. .16136	/rpt_family="ERV1"

repeat_region /rpt_family="ERV1"
16178..16239 /rpt_family="ERV1"
16728..17201 /rpt_family="ERV1"
17188..17253 /rpt_family="ERV1"
17243..17602 /rpt_family="ERV1"
17748..18064 /rpt_family="ERV1"
18085..18235 /rpt_family="L1"
18303..18876 /rpt_family="ERVK"
18927..19413 /rpt_family="ERVK"
19419..19628 /rpt_family="ERVK"
21508..21759 /rpt_family="L1"
22725..22868 /rpt_family="Alu"
22868..23335 /rpt_family="ERV1"
23350..24076 /rpt_family="ERV1"
24273..24690 /rpt_family="ERV1"
24702..24823 /rpt_family="ERV1"
24886..25175 /rpt_family="Alu"
25186..25753 /rpt_family="ERV1"
27654..27697 /rpt_family="ERV1"
28186..28421 /rpt_family="ERV1"
28440..28804 /rpt_family="ERV1"
28773..29267 /rpt_family="ERV1"
29270..31939 /rpt_family="ERV1"
31933..32211 /rpt_family="ERV1"
32225..32546 /rpt_family="Alu"
32582..33624 /rpt_family="ERV1"
33740..34049 /rpt_family="ERV1"
34050..34149 /rpt_family="ERV1"
34155..34617 /rpt_family="ERV1"
34624..37556 /rpt_family="ERV1"
37676..37979 /rpt_family="ERV1"
38072..38379 /rpt_family="ERV1"
38380..39312 /rpt_family="ERV1"
39318..40245 /rpt_family="ERV1"
40372..40443 /rpt_family="ERV1"
40486..41009 /rpt_family="ERV1"

Query Match

15.5%; Score 215.6; DB 9; Length 161879;

Best Local Similarity 62.1%; Pred. No. 3.2e-50;
Matches 390; Conservative 0; Mismatches 234; Indels 4; Gaps 3;
Qy 642 AGTTGTGCTAATGGTATTTTCCTCGGTATCTGTCACTATGGGGCCACCTTTTGGCAATAT 701
Db 160312 AGCATTTGCTGATGACATGTTCTTCTTACCACTTGTGTATGGGGCAATTCACATGCAATAC 160371
Qy 702 GTTGGCTATCCAGATCAAGTACGATGACTACAAGATTTCGCCCTTGGGCCACTAGAACTCCT 761
Db 160372 ACTGGCCATCCATACCAGCAAGCACTTTCCTGACATTAAGCTACAGCGCATAAATTCAT 160431
Qy 762 CTGCATCAACCATCTGGCGGACATTTGGAGATCACTTCCCGCCCTCTCGAATTCCTGGTCTCTT 821
Db 160432 CTG--TCATGATACGGTGTCTCTTGGTGATTAATCTCACGTGTAGTACCTCTGGAATTTT 160489
Qy 822 CTCAGCCACTTTGAAATTTGAAGGCTGTGCCCTTCTCTAGTGTCAACTCTCTGATCATCTCT 881
Db 160490 TCCTGCATCTCTGAAACAGAGGAGCCTACCAATTAATCATATATATTTTGTATTATT 160549
Qy 882 CTTTGAGCCCTGGATTAAAGTTCTCGAGAAAGTGTCGCCAGATGCCCAATAACATTTCAGAA 941
Db 160550 GTTGCACCATGGCTGGAGTTTGGAAAGTGGAGCTCATCTTCATAACACACAGAAAA 160609
Qy 942 AAATTCAGCCGGGTGGCACTCTGGTGGTCTCTGATTTTCAGTCACCATCTCTATCTCTG 1001
Db 160610 TAAATTCAGCATGGTGGGTAC-ATACGGATGCTTACCTTAATCATGCTACTATATCTGTC 160668
Qy 1002 CATCAACTTCTCTGCTGGTGCAGCTTTTCGAGTTGAGTTGGCAGACAGAGATCTCGTCGA 1061
Db 160669 TATCAACTTCTCTGGCTGGTCAGCACTGAAACTGCAGCTCTCAAAATGAGGAAGTAATTGA 160728
Qy 1062 CAAAGGGCAGAACTGGGGCAGATATGGGCTGCACATATAGTGTGAGGTTGGTAGAGATGT 1121
Db 160729 CAAGAGACTGAGGTGGGC-CATAGAAATCCTACACTACAGCTTCTAGTTTTTAAAGAAATGT 160787
Qy 1122 GATCATGCTCTTGGTTTTTAAAGTTCTTTGGAGTGAAGTGTGTACTGAAATTAATCTGTCAATTC 1181
Db 160788 GATAATAATATTGATATTATTTATGTTTCGTTGGAGGGAATAATTTCACTGAAGTCTTTGTGACTC 160847
Qy 1182 CTGATTCGCCCTTGCAGCTCATTTATGCTTTATCTGATTTTCCATTTGGCTTCATGCTCTCTTTT 1241
Db 160848 ATTAATTCGCATGTAGTTTCATCAAACTAACTAATTAGCCATTGGCTTTATGCTCTCTCTT 160907
Qy 1242 CTTCCAGTACTTGCATCCATTGGCGCTCA 1269
Db 160908 CTGTCAGTATTTCACCAAGGTGGTCA 160935

Search completed: March 30, 2003, 05:28:41
Job time : 3999.31 secs